DOCUMENT RESUME

ED 168 055 CS 204 769

TITLE Testing and Evaluation in English and the Language Arts: Abstracts of Doctoral Dissertations Published in "Dissertation Abstracts International," January through December 1978 (Vol. 38 No. 7 through Vol. 39 No. 6).

INSTITUTION ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill.

PUB DATE 78

NOTE 11p.; Parts may be marginally legible due to light print

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Articulation (Speech); Communication Skills; Criterion Referenced Tests; Doctoral Theses; Elementary Secondary Education; *English Instruction; *Evaluation Methods; Grading; Kindergarten Children; *Language Arts; Listening Comprehension; Oral English; Story Reading; *Student Evaluation; *Testing; Writing Skills

ABSTRACT

This collection of abstracts is part of a continuing series providing information on recent doctoral dissertations. The 11 titles deal with the following topics: interrelationships among different approaches to measuring listening comprehension; two methods of articulatory assessment on the responses of kindergarten children; a comparative appraisal of two language arts testing procedures; influences on the evaluators of student writing; assessing factors in children's writing to determine their level of sophistication; first grade children's understanding of the oral language of familiar and unfamiliar adults; the consistency of the Scholastic Aptitude Test--Verbal Section (SAT-V) with present objectives in high school English; the construction and validation of a criterion-referenced test in basic writing skills; the effects of readers' fatigue on the grading of essays; a comparative study of the effects of illustrations in the testing of language comprehension in children; and the effect of three different methods of story presentation on the language performance of third grade students. (MAI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED168055

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Testing and Evaluation in English and the Language Arts:

Abstracts of Doctoral Dissertations Published in Dissertation Abstracts International, January through December 1978 (Vol. 38 No. 7 through Vol. 39 No. 6)

Compiled by the Staff of

The ERIC Clearinghouse on Reading and Communication Skills

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

University Microfilms International

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

C 5204769

The dissertation titles contained here are published
with permission of the University Microfilms International,
publishers of Dissertation Abstracts International
(copyright © 1978 by University Microfilms International),
and may not be reproduced without their prior permission.

This bibliography has been compiled as part of a continuing series designed to make information on relevant dissertations available to users of the ERIC system. Monthly issues of Dissertation Abstracts International are reviewed in order to compile abstracts of dissertations on related topics, which thus become accessible in searches of the ERIC data base. Ordering information for the dissertations themselves is included at the end of the bibliography.

Abstracts of the following dissertations are included in this collection:

Binford, Francis Emerson Jr.

A STUDY OF INTERRELATIONSHIPS AMONG DIFFERENT APPROACHES TO MEASURING LISTENING COMPREHENSION

Brunner, Mary Catherine

EFFECTS OF TWO METHODS OF ARTICULATORY ASSESSMENT ON THE RESPONSES OF KINDERGARTEN CHILDREN

Butler, Roberta Jacqueline

A COMPARATIVE APPRAISAL OF THE UNIVERSITY OF GEORGIA SECONDARY ENGLISH EDUCATION EVALUATION PROCEDURES AND THE GEORGIA CRITERION-REFERENCED TEST IN COMMUNICATIVE ARTS

Freedman, Sarah Warshauer

INFLUENCES ON THE EVALUATORS OF STUDENT WRITING

Hohmann, Charles Frederick

ASSESSING THE EPISODIC STRUCTURE, INFORMATION STRUCTURE AND COHESION OF CHILDREN'S WRITTEN NARRATIVES AS INDICES OF THEIR SOPHISTICATION IN WRITING

Kear, Dennis Jay, Sr.

ORAL LANGUAGE ELICITATION BY FAMILIAR AND UNFAMILIAR ADULTS

Moran, Glenn Thomas

THE CONSISTENCY OF THE SAT-V WITH PRESENT OBJECTIVES IN HIGH SCHOOL ENGLISH

Prater, Doris Lee

THE CONSTRUCTION AND VALIDATION OF A CRITERION-REFERENCED TEST IN BASIC WRITING SKILLS

Steller, Nancy Ann Dodrill

THE EFFECTS OF READERS' FATIGUE ON THE GRADING OF ESSAYS

Straub, John Michael

A COMPARATIVE STUDY OF THE EFFECTS OF COLORED VS. BLACK AND WHITE ILLUSTRATIONS IN THE TESTING OF LANGUAGE COMPREHENSION IN CHILDREN

Walker-Dalhouse, Doris Jean

THE EFFECT OF THREE DIFFERENT METHODS OF STORY PRESENTATION ON THE LANGUAGE PERFORMANCE OF A SELECT GROUP OF THIRD GRADE STUDENTS

A STUDY OF INTERRELATIONSHIPS AMONG DIFFERENT APPROACHES TO MEASURING LISTENING COMPREHENSION

BINFORD, Francis Emerson Jr., Ph.D.
The University of Iowa, 1977

Supervisor: Professor Albert N. Hieronymus

This study was designed to investigate the interrelationships among different approaches to measuring listening comprehension among fifth and sixth grade students. The investigation had four primary goals. The first was to conduct a comprehensive review of the research on listening comprehension. The second was to construct a compendium of listening as a result of that initial review. The third was to construct an experimental test battery with subtests that defined eight selected comprehension skills. The fourth was to investigate the nature of relationships among the subtests in the experimental battery and to examine the relationships between listening and achievement measures of reading and composite basic skills performance.

The first two goals were met by a review of the listening research literature that found listening and reading comprehension were two highly related, but apparently separate abilities. Listening and reading were found to be taught and measured by similar methods. Then a compendium of twenty-two listening skills were found in the tests of listening comprehension, that appeared to be highly interrelated and overlapping.

This led to the construction of the experimental battery of eight subtests, that were divided into two half-test forms, A and B, and consisted of sixty-two items in each form. The two half-test forms were administered approximately one week apart.

The sample was selected from among those schools participating in the 1975, spring administration of the Iowa Basic Skills Testing Program and was composed of 396 students from four elementary schools, at the fifth and sixth grade levels. The students selected were from medium sized schools and consisted of a sample of fairly representative variability.

Equivalent half-forms reliability coefficients were computed for the eight subtests and total listening test. The subtest intercorrelations corrected for attenuation were computed utilizing the averaged half-length and full-length test intercorrelations and were tested for significant departure from $r = 1.00$.

Full-length test intercorrelations and reliability coefficients were used to compute reliabilities of differences among subjects

Two factor analysis solutions were performed on the data. First, the principal factor analysis procedure, followed by a varimax rotation of the initial solution was conducted for factors with eigenvalues greater than one. Second, the principal factor analysis procedure, followed by a varimax rotation was conducted for a two-factor solution.

Although performance on the majority of the listening tests favored girls, differences between means for girls and boys were not significant. However there were sizable and significant sex differences favoring girls in reading and on the ITBS composite.

The reliability coefficients, computed from split halves ranged from .36 to .53 for the eight subtests. The reliability coefficient for the total test was .84.

The corrected correlations, reliabilities of differences, and factor analysis procedures indicated little evidence of uniqueness among listening subtests. Few of the correlations corrected for attenuation among the listening subtests were significantly below 1.00 and reliabilities of differences were very low. Only one listening factor was found.

The results of this study lend little or no support to providing separate subtest scores for different facets or dimensions of listening comprehension, or for constructing a diagnostic battery of listening skills. A single general listening ability factor appeared to account for nearly all of the systematic variance. These results are consistent with similar research on reading comprehension.

Reading and listening comprehension are highly related, but significantly independent abilities with strong similarities in definition, development, and assessment.

Order No. 77-28,431, 262 pages.

EFFECTS OF TWO METHODS OF ARTICULATORY ASSESSMENT ON THE RESPONSES OF KINDERGARTEN CHILDREN

Order No. 7800386

BRUNNER, Mary Catherine, Ph.D. University of Maryland, 1977. 135pp. Supervisor: Dr. Sarah H. Leeper

The primary purpose of this study was to investigate the relationships among certain variables which might influence the articulation test scores of kindergarten children. The variables selected for study were method of response elicitation, examiners, and sex of subjects. Stimulus material consisted of 87 words from The Templin-Darley Tests of Articulation. Two test methods were established for presentation of the stimulus. Method One was a spontaneous method in which pictures of objects or concepts were presented to elicit a word containing given speech sounds. Method Two was an imitative method in which subjects were asked to repeat a word after hearing it pronounced by the examiner.

Randomly selected from five randomly selected elementary schools in Baltimore County, Maryland, 168 kindergarten children were subjects. This sample included equal numbers of boys and girls and consisted of children who met specific criteria. They manifested no evidence of hearing loss; evidenced no malformed or abnormally functioning speech production or sound reception organs; came from monolingual homes; were not involved in speech therapy; and received parental permission to participate in the study.

A Posttest Only Control Group Design was used to ascertain the effect of each of the two test methods upon articulatory responses of kindergarten children. The structure of the study resulted in a Mixed Model 2x2x2 factorial design, and the independent variables (test method, examiners, and sex of subjects) were treated in a Three-Way Analysis of Variance to test the statistical significance of both main effects and joint effects of those variables. The level of significance for testing all research questions was set at α .05.

Results of this study were: 1. Method of test administration did affect the subject's performance on a test of articulation. Difference in mean scores was significant at the .05 level in favor of the imitative method. 2. Test performance was related to the sex of the subject. Difference in mean scores was significant at the .05 level in favor of girls. 3. Analysis of data related to main effects of examiners produced no statistically significant results. However, data related to effects of interactions between test method and examiners indicated that examiners had an effect that was not independent of method.

Some supplementary findings of this study were: 1. Girls made fewer errors than boys in every error category, except that of omissions. (Error categories were: omissions, substitutions, distortions, and no response.) 2. Twice as many boys as girls received low scores on a test of articulation. 3. Ages of children who received perfect scores covered the total range of the sample – five years, two months to six years, one month – except for two age levels for boys (five years, four months and five years, eight months) and one age level for girls (five years, three months).

A COMPARATIVE APPRAISAL OF THE UNIVERSITY OF GEORGIA SECONDARY ENGLISH EDUCATION EVALUATION PROCEDURES AND THE GEORGIA CRITERION-REFERENCED TEST IN COMMUNICATIVE ARTS

Order No. 7822296

BUTLER, Roberta Jacqueline, Ed.D. University of Georgia, 1978. 166pp. Supervisor: L. Ramon Veal

This study examined the traditional method of evaluating teachers with special emphasis on the directions suggested by reviews of research and of related legal issues. These reviews indicate that a multiple evidence approach to teacher evaluation seems the safest direction. Both the University of Georgia's field-based Pre-Service Secondary English Education Program and the Georgia Department of Education have adopted such an approach. The Georgia Department of Education plans to utilize one unique piece of evidence, scores on a knowledge test, the Georgia Teaching Field Criterion-Referenced Test in Communicative Arts (CRT-CA), in addition to other certification requirements.

There were two purposes to this study:

1. To further validate the instruments used by the University of Georgia Pre-Service Secondary English Education Program to measure prospective English teachers' knowledge, attitudes, and performance.
2. To use these instruments in a comparative, evaluative study of the Georgia Teaching Field CRT-CA.

The population for the study consisted of 40 teacher candidates (TC's) who completed the University of Georgia program in 1975-1976 and who were involved in a previous study which evaluated the program, and, in addition, 40 TC's who completed the program in 1976-1977 and who took the Georgia Teaching Field CRT-CA as part of a large scale field tryout of the test.

A comparison of these two groups' scores on the University of Georgia instruments, which measure 26 variables, established their similarity and provided further evidence of the reliability of the instruments. Principal component analyses of the scores of all 80 students on the 26 variables extracted two knowledge factors (Tested Knowledge and Academic Success), three attitude factors (Social Orientation, Acceptance of Responsibility and Leadership, and Beliefs about Students and Teaching), and two performance factors (Professional Ratings and Student Ratings). The limited number and logical clustering of these factors further established the validity of the instruments. The principal component analyses also produced factor scores thereby reducing the number of variables to be included in the correlational analyses which followed.

Using the 40 1976-1977 TC's scores on the seven factors and on the CRT-CA, an intercorrelation matrix revealed only three significant correlations among the eight variables. There is a positive relationship between the two knowledge factors (Tested Knowledge and Academic Success) and the CRT-CA and a negative relationship between Tested Knowledge and Professional Ratings. Stepwise multiple regressions reiterated these relationships and indicated that a combination of Tested Knowledge and Academic Success scores provides the best prediction of the CRT-CA score but that no combination of scores provides prediction of Professional Ratings. In sum the CRT-CA does not predict performance as measured by Professional Ratings, nor do Professional Ratings or Student Ratings of performance predict scores on the CRT-CA. The study revealed no significant relationship between the CRT-CA and any of these measures of performance. The results force the conclusions that none of the predictor measures holds much potential for positive prediction of classroom performance and that these measures of knowledge hold little potential for predicting anything other than performance on other instruments also measuring knowledge.

INFLUENCES ON THE EVALUATORS OF STUDENT WRITING

Order No. 7802159

FREEDMAN, Sarah Warshauer, Ph.D. Stanford University, 1977. 230pp.

This thesis asked why competent evaluators award the ratings they do to college students' expository essays. Most past studies on composition evaluation gave only partial answers to this question because they were based on correlational research designs. Unlike past studies the main section of this thesis consisted of an experiment designed to discover factors within papers that influence raters.

Essays were collected from college students on eight argumentative topics. Some of these essays were rewritten to be stronger or weaker in four categories: content, organization, sentence structure, and mechanics. Evaluators, not knowing that the essays had been rewritten, first used a four point holistic rating scale to judge the quality of the essays. Then they rated whether they thought each of the four rewriting categories in each rewritten essay was strong or weak. These ratings of the categories were labeled the raters' perceptions of the rewriting.

An analysis of how the rewritten categories influenced the variance in the holistic scores revealed that content and organization were the most influential ($p < .001$). The effects of the mechanics and sentence structure were smaller and differed when measured by the actual rewriting and when measured by the evaluators' perceptions of the rewriting. Mechanics and sentence structure mattered more when organization was strong ($p < .01$, $p < .001$, respectively).

Examination of the original essays prior to rewriting revealed effects of the writer, the rating scale, and the circumstances of the rating. Essays from a group of professional writers were also in this sample. All original essays were evaluated with both a holistic and a Diederich type analytic rating scale.

Two writer attributes contributed significantly to the raters' holistic evaluations of the quality of the original essays: where the writer went to school ($p < .001$) and how confident the writer was in his or her in-class writing ($p < .05$). The trainer ($p < .01$), the essays themselves ($p < .001$), and one of the topics significantly influenced the raters' holistic scores. Neither the time of the reading nor the raters themselves influenced the scores to differ significantly.

The professional writers fared only slightly better than the average student on the holistic scale though they did much better than the students on the analytic scale. Their outstanding analytic scores came from the categories voice, sentence structure, word choice, and usage. They were not judged so consistently high on the categories development and organization.

Students usually received the same scores regardless of the rating scale. The summed score on an analytic rating scale, the holistic score, and every category of the analytic scale, except usage, were highly correlated.

This thesis provides an experimental model for studying many aspects of the evaluation process. It suggests guidelines for obtaining consistent evaluations from raters and for analyzing what those ratings mean. It also points to the need for both pedagogy and research to focus attention on the larger levels of the discourse, content and organization.

ASSESSING THE EPISODIC STRUCTURE, INFORMATION STRUCTURE AND COHESION OF CHILDREN'S WRITTEN NARRATIVES AS INDICES OF THEIR SOPHISTICATION IN WRITING Order No. 7822917

HOHMANN, Charles Frederick, Ph.D. The University of Michigan, 1978. 144pp. Chairman: Loren Barritt

In an effort to extend recent work at the High/Scope Educational Research Foundation, ninety narrative stories written by third, fourth and fifth grade children in an administration of the Foundation's Productive Language Assessment Tasks (PLAT) were scored on variables representing episodic structure, information structure and cohesion. The variables were based on models of episodic structure, information structure and cohesion identified in a review of relevant literature and involved the tabulation of word, phrase and larger story constituents.

The episodic structure, information structure and cohesion variables showed development changes in most cases where expected. In addition the variables showed significant relationships to analogous PLAT variables and to global ratings of the stories' quality. Better quality stories were long, ended with a resolution, contained several episodes, used the causal conjunction so frequently and had few distracting errors. These and other findings are potentially applicable to writing instruction and assessment.

ORAL LANGUAGE ELICITATION BY FAMILIAR AND UNFAMILIAR ADULTS Order No. 7820684

KEAR, Dennis Jay, Sr., Ph.D. Arizona State University, 1978. 91pp.

The purpose of this study was to compare the syntactic structures produced by first grade children using a specific oral language elicitation instrument and how the degree of examiner familiarity effected their oral language performance. The variables of socioeconomic status and ethnic group membership and their effect on oral language production were also examined. The Oral Language Evaluation (O.L.E.) was chosen to elicit and evaluate the oral language responses of 68 first grade Mexican-American and Anglo children. The middle and low socioeconomic groups were included in the study. Three different types of examiner were utilized to elicit the oral language responses--a familiar adult, a classroom teacher and an outside evaluator. A three-way ANOVA was used to analyze the data. No statistically significant differences were found in the oral language scores on the O.L.E.'s language continuum between socioeconomic groups, ethnic groups, and types of examiners. Two and three-way interactions among the three independent variables also proved to be statistically not significant.

THE CONSISTENCY OF THE SAT-V WITH PRESENT OBJECTIVES IN HIGH SCHOOL ENGLISH Order No. 7730851

MORAN, Glenn Thomas, Ed.D. University of Northern Colorado, 1977. 116pp.

This study was designed to investigate the following questions: 1. Are the present educational programs in high school English teaching objectives consistent with the objectives tested for by the Scholastic Aptitude Test - Verbal Section (SAT-V)? 2. Should our high school English programs be consistent with the objectives tested for by the SAT-V?

The population of the study consisted of high school English chairpersons. They responded to questionnaires sent to randomly selected high schools in Colorado. The data and conclusions are based on the returns representing 41 high schools.

The questionnaire was composed of sample questions from the four areas found on the SAT-V. These areas were: sentence completion, analogies, antonyms and reading comprehension. The selected participants by means of a Likert scale, indicated whether or not each question was consistent with the goals of their educational programs and whether or not they believed their program objectives should be consistent with the objective tested for by the SAT-V.

A descriptive analysis was made of each question according to the following variables: sex, years taught, size of high school, and the percentage of minority students within each high school.

Some of the conclusions drawn were: 1. Most of the high schools sampled felt their educational programs were, and should be, designed to teach the objectives tested over by the SATs. 2. Although the sample had mixed feelings as to the SATs being a good indicator of education experiences, they did feel that the SATs should continue to be given. 3. The sample of English teachers felt that a student's acceptance to college not be determined by SAT scores.

The review of literature and data collected resulted in the following recommendations being made: 1. Educators and parents should be made aware of the other variables involved in the declining SAT scores. 2. Institutions of higher education, parents, and students should be made aware that not all high schools have educational objectives that are consistent with the objectives tested by the SAT-V. 3. Similar studies in other states should be made to see if the results are consistent with this study. 4. College success of students who graduated from high schools which did not support the objectives of the SAT-V is a recommendation for further study. 5. In respect to the research on declining SAT scores and with the consideration of some of the conclusions, it should be recommended that the SATs be discontinued.

THE CONSTRUCTION AND VALIDATION OF A CRITERION-REFERENCED TEST IN BASIC WRITING SKILLS
Order No. 7811372

PRATER, Doris Lee, Ed.D. University of California, Los Angeles, 1978. 197pp. Chairman: Professor W. James Popham

The need for the development of well constructed criterion-referenced tests of basic writing skills has been enhanced by the current trend toward establishing minimum requirements in basic skill areas for high school graduation. This study documents a replicable procedure for the identification of basic writing competencies, construction of test specifications and items, and establishment of the validity and reliability of such a measure.

The first step in the procedure was to gather preference data regarding potential competencies from several involved subgroups within a school district. A mailed survey listed a wide range of potential writing competencies and asked respondents to prioritize them as to their importance as a requirement for high school graduation. Survey forms were mailed to parents, teachers, students, business representatives, and selected subject matter specialists. On the basis of the preference data collected, seven competencies were isolated to form the basis for the test. These competencies were: Mechanics, Usage, Paragraph Organization, Spelling, Sentence Correction, Forms, and Writing Sample.

The second step in the procedure was the creation of test specifications for each of the competencies selected. The specifications included a brief general description of the competency, a sample item, and a list of both stimulus and response attributes. Using the test specifications, twenty multiple-choice items were generated for six of the competencies. Ten of the items for each of these competencies were assigned to Form A of the test, ten to Form B. The seventh competency, "Writing Sample", required a constructed response on the part of the examinee. For this subtest, a scoring guide, or rubric, was constructed and English teachers were trained to use the guide in scoring essays.

The final step in the procedure was the establishment of the validity and the reliability of the test. The descriptive validity of the test was established by two procedures. First, in order to determine the extent to which the test specifications communicated clearly, four experienced language arts teachers were given the test specifications for each competency and asked to create, independently, three test items according to the specifications for that competency. In order to check the homogeneity of the items produced by these language arts teachers, three other language arts specialists were given the items and asked to identify those that were homogeneous. A second validation procedure involved the actual test items. Five judges reviewed the items that were used in the test and judged whether or not these items were congruent with the test specifications. The test was field tested on three hundred students randomly selected from grades seven, nine, and twelve. Two weeks later, the same students were retested. The following types of reliability were calculated for the total test and each subtest: Stability, Equivalency, and Stability and Equivalency. Results from field tests yielded correlation coefficients of .88 and .86 on test-retest results (stability), .92 and .93 on alternate forms results (equivalency); and .84 and .87 on the basis of results across time and form (stability and equivalency).

In addition, consistency of classification of students as either "Pass" or "Fail" was computed using varying proficiency levels of seventy, eighty, and ninety percent. Consistency of decisions ranged from seventy-eight to one hundred percent, depending upon the proficiency level chosen. Inter-rater agreement of readers using the essay scoring guide ranged from a low of .73, after the first administration, to a high of .93 after the second administration.

THE EFFECTS OF READERS' FATIGUE ON THE GRADING OF ESSAYS
Order No. 7822356

STELLER, Nancy Ann Dodrill, Ph.D. University of Georgia, 1978. 112pp. Supervisors: Dr. George E. Mason and Dr. R. Robert Rentz

Statement of the Problem

The purpose of this study was to investigate the effects of readers' fatigue on the graders of written essays. The essays used for the study were written for the Language Skills Examination of the Regents' Testing Program. The focus of the study concerned the variability of reader judgments as they might relate to the effects of readers' fatigue over time of grading. Three hypotheses, tested at the .05 level of significance, were generated.

Hypothesis One

There will be no significant difference between means for experienced and inexperienced graders.

Hypothesis Two

There will be no significant difference between means for fatigued and non-fatigued graders.

Hypothesis Three

There will be no significant interaction between experience and fatigue.

Method and Procedure

The sample was a random selection of graders who read essays during the time period from Winter Quarter, 1975, through Summer Quarter, 1976. Before random selection, raters were designated as "experienced" (five or more grading days within the seven inclusive quarters), or "inexperienced" (three or fewer grading days within the seven inclusive quarters). Of the raters whose reading averaged 110 to 150 essays per scoring day, forty in each category, experienced and inexperienced, were randomly chosen. The random sample was independent of the six scoring centers located within the University System, although the sample did consist of at least two raters in both categories from each center.

The basic technique used in the study was a comparison of "first-half" grades with "last-half" grades, determined from the score sheets used by the readers. The data consisted of 160 sets of scores, with a minimum of 55 scores per rater per half.

Assumption of the presence of more readers' fatigue in the last half of the grading time than in the first half constituted the operational definition of the independent variable "fatigue." As this assumption suggests the possibility of more variation in scores given in the last half of the grading time than in the first half, variation in fatigue was observed by comparing first-half grades with last-half grades. This possible variation, then, constituted the operational definition of the independent variable "fatigue." A two-way analysis of variance was used to analyze the data.

Results

Examination of the results of the analyses indicated that the null hypotheses could not be rejected.

1. The F value of .4509 (experienced and inexperienced graders, main effect (E)) was not significant at the .05 level.
2. The F value of .0217 (fatigued and non-fatigued graders, main effect (F)) was not significant at the .05 level.
3. The F value of .0491 (interaction between experience and fatigue, (FxE)) was not significant at the .05 level.

Conclusions

The obvious conclusion of this investigation is that, within the limitations of the study and according to the interpretation of the data gathered for the study, there is no significant effect of readers' fatigue on the grading of essays.

A COMPARATIVE STUDY OF THE EFFECTS OF COLORED VS. BLACK AND WHITE ILLUSTRATIONS IN THE TESTING OF LANGUAGE COMPREHENSION IN CHILDREN
Order No. 7821774

STRAUB, John Michael, Ph.D. Indiana University, 1978. 224pp. Chairperson: Dr. Ivor K. Davies

The purpose of this study was to compare the performance of both language disordered and normal children on two versions of a language comprehension task: one task illustrated with simple black and white line drawings, and the other illustrated with the same drawings filled in with color. Past research in perceptual discrimination and dimensional preferences have indicated that children younger than age seven may have difficulty interpreting the intended meaning of simple black and white line drawings, but would have less difficulty if the same drawings were rendered in color.

An experimental language comprehension task was developed based on Lee's Northwestern Syntax Screening Test, Receptive Portion (1969). Two versions of the experimental task were prepared, one containing black and white line drawings, and the other containing the same drawings filled in with color. Twenty language disordered children and twenty normal language children between the ages of three and ten were designated according to age level and language ability.

All of the children were given both versions of the experimental task. A one week time interval was allowed between administrations. Five Independent Variables were manipulated, each with two levels: Chroma (Color Vs. Black and White), Status (Language Disordered Vs. Normal), Age (Younger Vs. Older), Level (High Language Ability Vs. Low Language Ability), and Order (Color Task First Vs. Black and White Task First). The Dependent Variable was each child's score on the experimental language comprehension task.

The analysis of variance yielded two statistically significant main effects, Status and Age, and two significant statistical interactions, Chroma X Order and Age X Order. All were significant at the .01 level of significance. The results demonstrated that a significant learning effect occurred between the two administrations of the experimental task causing the lack of significant differences between the two levels of Chroma for all the children.

Despite the learning effect, significant differences between the two age levels of children in regard to the presence or absence of color emerged as a result of counterbalancing the order of presentation of the two task conditions. The younger children performed significantly better on both conditions when given the Color Task first rather than the Black and White Task. The older children, conversely, performed significantly better on both conditions when presented the Black and White Task first rather than the Color Task ($p < .01$). Both language disordered and normal groups consistently followed this direction, but the greatest difference in performance between the age groups was among the language disordered children.

The results of this study indicate that language disordered children under the approximate age of seven are only able to perform adequately on language comprehension tests when the illustrations are rendered in color. The results also indicate that language disordered children above the age of seven are distracted by the presence of color in the test illustrations, and perform to their potential only when the illustrations are in black and white.

THE EFFECT OF THREE DIFFERENT METHODS OF STORY PRESENTATION ON THE LANGUAGE PERFORMANCE OF A SELECT GROUP OF THIRD GRADE STUDENTS
Order No. 7731848

WALKER-DALHOUSE, Doris Jean, Ph.D. The Ohio State University, 1977. 144pp. Adviser: Professor James Kerber

The purpose of this study was to determine if different methods of story presentation would enhance or decrease language performance (total number of words or fluency, excluding garbles; vocabulary sophistication or word frequency; and language complexity), and to determine if any of these story telling procedures interact with the student's sex, race, and educational background of his/her parents.

Forty-eight (48) subjects were randomly selected from the population of third grade students in the Columbus, Ohio Public School System. Of the forty-eight (48) students, twenty-four (24) were male and twenty-four (24) were female. Of the male and female groups of subjects selected, twelve (12) were black and twelve (12) were white. Six (6) or one-half of each group were from homes of professional class parents, and six (6) were from homes of non-professional class parents.

Three methods of story presentation were used in the study. Each subject received all three methods of story presentation. Treatment I was a pictures only (P) presentation of a story. Treatment II was a pictures and recording (P + R) presentation of another story. Treatment III was a recording (R) of a third story. The three stories used in the treatments were selected from the Caldecott Award Winner Picture Books. After each story presentation, the subjects were asked to retell or tell the story back to the investigator.

It was observed that when third grade students were presented stories by Treatment III, recording only (R) and Treatment II, pictures and recording (P + R), there was a significant effect on their language performance as measured by the

mean number of words produced or fluency, excluding garbles, and language complexity (excluding words per clause). Treatment I, pictures only (P), produced more words per clause for all categories of subjects.

A significant main effect was found for the educational background of students' parents on the mean number of words used in story retelling. A significant main effect was also found for race on vocabulary sophistication or word frequency. Analysis of the data also revealed a significant interaction effect between sex and race for vocabulary sophistication.

The data indicates that the auditory mode appears to be the most effective medium for enhancing language performance (as defined in the present investigation) for third grade students designated by sex, race (black and white), and educational background (professional and non-professional) of students' parents. Presenting stories in both the auditory and visual modes appears to be the next effective medium, while the visual mode appears to be the least effective in enhancing language performance.

Copies of the dissertations may be obtained by addressing your request to:

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106

or by telephoning (toll-free) 1-800-521-3042